Analysis of Recognition System of Japanese Sign Language using 3D Image Sensor

Yanhua Sun *, Noriaki Kuwahara**, Kazunari Morimoto ***

* oo_alison@hotmail.com
** noriaki.kuwahara@gmail.com
***morix119@gmail.com

Abstract: In this paper, we propose to introduce a recognition system of Japanese sign language using 3D image sensor for supporting the people who are deaf and mute or hard of hearing and speaking. We propose and implement a novel method for recognizing hand gestures using RGB and depth data from Microsoft's Kinect sensor. Beside of full body motions, our approach involves looking at specific hand gestures, to assist in the recognition of more refined gestures. With this approach, we are able to recognize 50 hand gestures of 52 hand gestures. This paper covers a new method to get the datasets of the hand gesture. However there are a lot of noises left so that the datasets cannot be used as it is, so PCL (point cloud library) effectively solved this optimization problem.

Key words: Japanese Sign Language , Kinect, recognition of JSL, ３D Sensor

1. Introduction

According to the investigation of 2006 about the population of people with disabilities of Japan, there are about 2.7million people, who are deaf and mute or hard of hearing and speaking. It just has 3 thousand people who have a job as a sign language translator. There are so little people to support the deaf and mute. In general life, people take the communication with the others through talking, but it is impossible for the deaf mute people, who can take the communication with others just using sign language. So it is necessary to explore a recognition system of sign language, which can help them to keep touch with others.

So the purpose of this paper is to develop a recognition system of Japanese sign language. About the system, Sign language recognition is such a system which translate the picture of the sign language to data, and then translate the data to words, sentences or sounds. This paper presents the front part of the recognition system. Several ways are used to research the sign language recognition, such as Facial movement analysis in ASL [7], human detection [3], 3D model tracking [4] and so on. But in most of the researches, even though they carry important grammatical and prosodic information, the rate of the recognition is low. We think the biggest problem is the first step: the way of taking hand gestures is not correct. About the 3D data, we just need the data of the finger, but the data of the whole environment is captured. It is hard to just take the hand data from the picture without any noises, and the part of the finger which should be recognized as 3D data sometimes is processed as noise by mistake, the reason like that make the rate of the recognition low.

For increasing the rate of the recognition, we think that the most important step is taking a 3D data with the least noise. First I want to introduce a 3D camera called Kinect. It is introduced by the company of Microsoft last year. Actually it is used to play the 3d game, and here, it is used as a 3D camera sensor. We found that if we use just one Kinect, we only can take the front data of hand; obviously the data of hand lost so much. The data we got is little, and it will directly affect the result of next step. So we think whether there is another way to get the 3D data as much as we can. As a result, we decided to use two Kinects to take the 3D data. Each Kinect can take a part of hand, and if we can do a data combination, so that we can take more data of hand.

The next, we had to discuss the direction of the Kinect, for example what kind of way the Kinect will be put, what kind of way the 3D data will be combined, through what kind of experiment to test the way, and so on. First we let two Kinects incline 90 degrees in two sides; we can get the front data of hand as much as we can. Through setting the parameter of the hand, we can get two kinds of data, the ones which are in the range of parameter, and the other ones which are out of it. The result of experiment shows that the ones in the range of parameter had a less error than the others which are out of range. It means that the way of combination is useful.

The second step, we plan to do a 3D data processing with PCL. PCL is an abbreviation of Point Cloud Library. The Point Cloud Library (PCL) is a standalone, large scale, open project for 2D/3D image and point cloud processing. Although the 3d sensor can take the important part more easily than before, there still have many noises. For example outliers, the noise from the environment, deviation value and so on. PCL supports natively the OpenNI 3D interfaces, and can thus acquire and process data from devices such as the PrimeSensor 3D cameras, the Microsoft Kinect and so on. So if I use the PCL in this research, I can do a perfect combination between the Kinect and the PCL. Use the advantage of the PCL, it not only can take the noise away, but also can reconstruct the surface of the hand, in some means, it can increase the rate of the recognition.

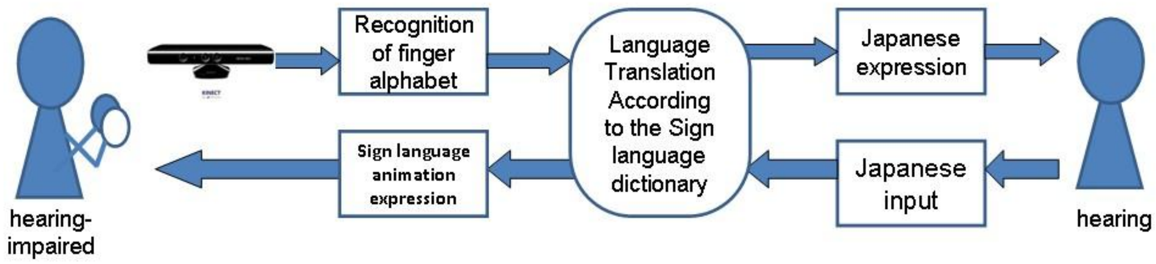

Figure.1 Recognition system of finger gesture

2. Condition and Method

2.1 Overview of method

This section provides an overview of the major steps in our method, which is summarized in Fig. 1. Implementation details are presented in Section 3 and Section 4. We first get two groups datasets of hand gesture, and do a combination through algorithm, reducing noise and smooth the datasets for later process. We use a 3D sensor to locate the hand gesture. In one side we first explore an environment to take the hand gesture only as much as possible, and in another side we must avoid taking the other parts except the hand gesture. So we extract the parameters of the head from the depth array and use the parameter to set the position of hand so that just the data of hand part can be gotten. Then we use the project we programmed to do a data set processing and the method called PCL. In this paper we just show the front part of the recognition system of Japanese sign language, and in next year, the part of datasets extraction, analysis and evaluation will be researched.

2.2 About Kinect

Launched in November 2010, Kinect is a motion sensing USB (Universal Serial Bus) input device by Microsoft that enables users to control and naturally interact with games and other programs without the need to physically touch a game controller or object of any kind. Kinect achieves this through a natural user interface by tracking the user's body movement and by using gestures and spoken commands [1, 2, 5]. Kinect holds the Guinness World Record as the fastest selling consumer electronics device, with sales surpassing 10 million units as of 9 March 2011. Kinect uses technology by Israeli company Prime Sense that generates real-time depth, color and audio data of the living room scene. Kinect works in all room lighting conditions, whether in complete darkness or in a fully lit room, and does not require the user to wear or hold anything. Prime Sense also teamed up with ASUS to develop a PC-compatible device similar to Kinect, which they called ASUS Xtion and launched in the second quarter of 2011. Running proprietary firmware (internal device software), these components together can provide full-body 3-D option capture, gesture recognition, facial recognition, and voice recognition capabilities.

Figure.2 Kinect Xtion

2.3 Research method

As it had been introduced, the Kinect is a strong powerful sensor, but also has weakness. Certainly the front part of hand which captured by Kinect can be taken, however the data of after part cannot be gotten. So in this part, we propose a novel method to solve the problem. It just showed as figure 3, and makes them perpendicular to each other, putting hand in the position where is scope of the parameter.

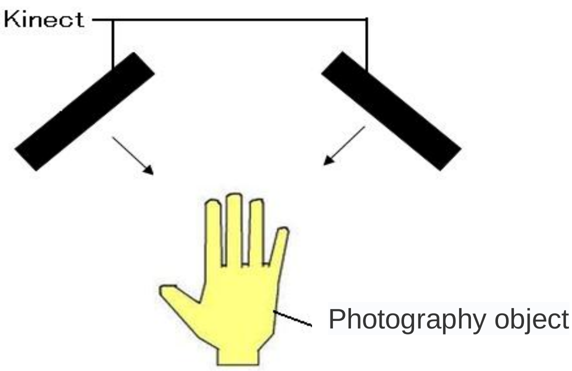

Figure.3 Method of using Kinect

Each Kinect images different parts, we use affine transformation algorithm (1) to combine the two datasets. In the algorithm C1 present the datasets of the left camera, C2 presents the datasets of the right camera showed as algorithm (2) . And we use transformation matrix algorithm (3) to do a data set transform of the cameras. After combination, we can know the accuracy of the superposition as the figure 4, about the result, obviously we can know that more part of hand had been taken.

$$X \cdot C_1 = C_2 \Longrightarrow X = C2 \cdot C1\text{-}1 \qquad (1)$$

$$\boldsymbol{C}_1,\ \boldsymbol{C}_2 = \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} \qquad (2) \qquad\qquad \mathbf{X} = \begin{bmatrix} a & b & c & tx \\ d & e & f & ty \\ g & h & i & tz \\ 0 & 0 & 0 & 1 \end{bmatrix} \qquad (3)$$

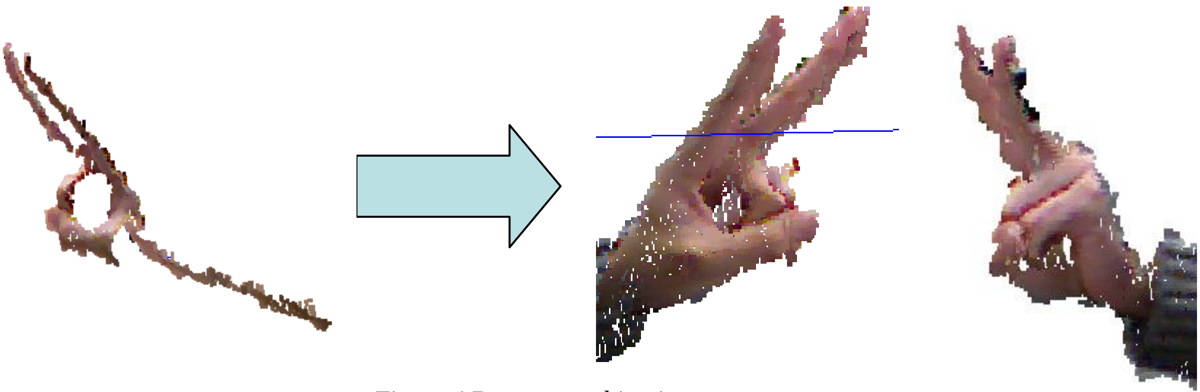

Figure.4 Dataset combination

4. Experimental Results

4.1 Evaluation of superimpose

The red circle express the measuring range, the blue one express all space, the green one express the range of parameter. Smaller the circles are, more accurate the positions are. So from the picture, we can know that the points which are within the scope of parameter are smaller than the points which are out of the scope. However it also proved that there all are a lot of noises which need to be cancelled. Then the method which will be introduced next resolve the problem.

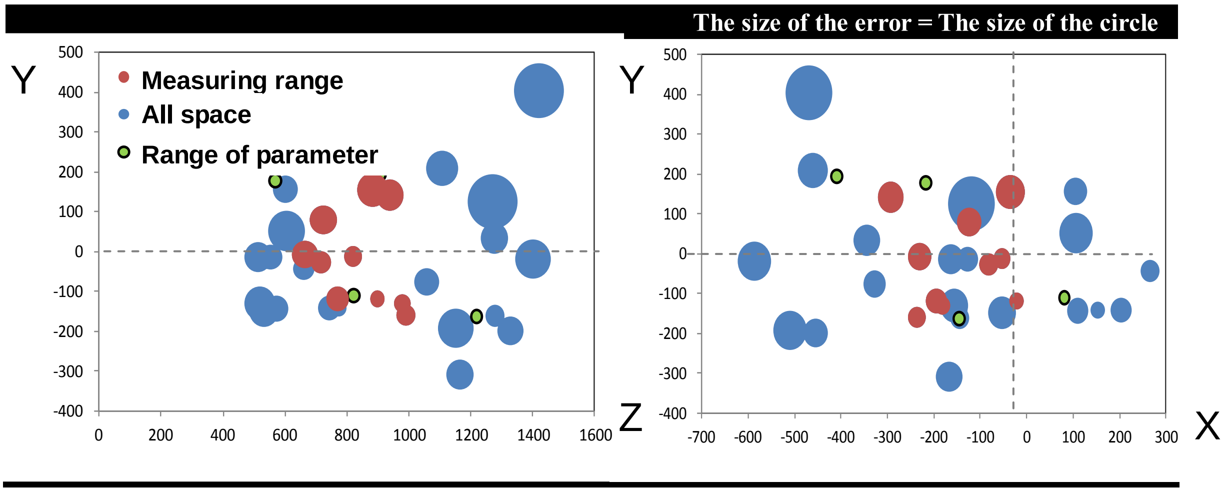

Figure.4 Recognition system of finger gesture

4.2 Evaluation of optimization

In this part, we will present some information about PCL (Point Cloud Library); PCL is an abbreviation of Point Cloud Library. The Point Cloud Library (PCL) is a standalone, large scale, open project for 2D/3D image and point cloud processing. So specifically what kind of 3D processing PCL can supply? Although the 3d sensor can take the important part more easily than before, there still have many noises. For example outliers, the noise from the environment, deviation value and so on. PCL supports natively the OpenNI 3D interfaces, and can thus acquire and process data from devices such as the Prime Sensor 3D cameras, the Microsoft Kinect and so on. The second step is to process the 3D data with PCL.

In this program, note that the coordinate axes are represented as red (x), green (y), and blue (z). The points are represented with green as the points remaining after filtering and red as the points that have been removed by the filter. In the figure 5, the left figure is the datasets which are before the PCL 3D data processing, and the right one is the datasets which are after the PCL 3D data processing, obviously we can know that most of red points which should be removed had been removed.

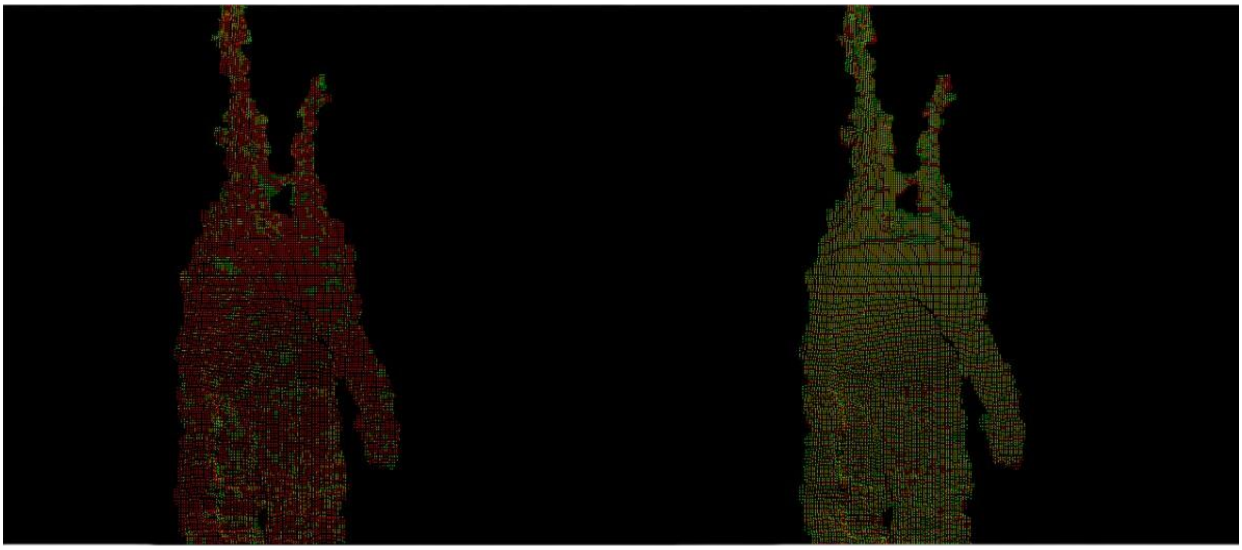

Figure.5 Before and after

Just as the figure 6 showed, the realistically environment is very complicated, and in the cloud viewer windows we can clearly see just the part of hand is captured perfectly, it completely proved that the method of taking datasets and 3D processing are successful.

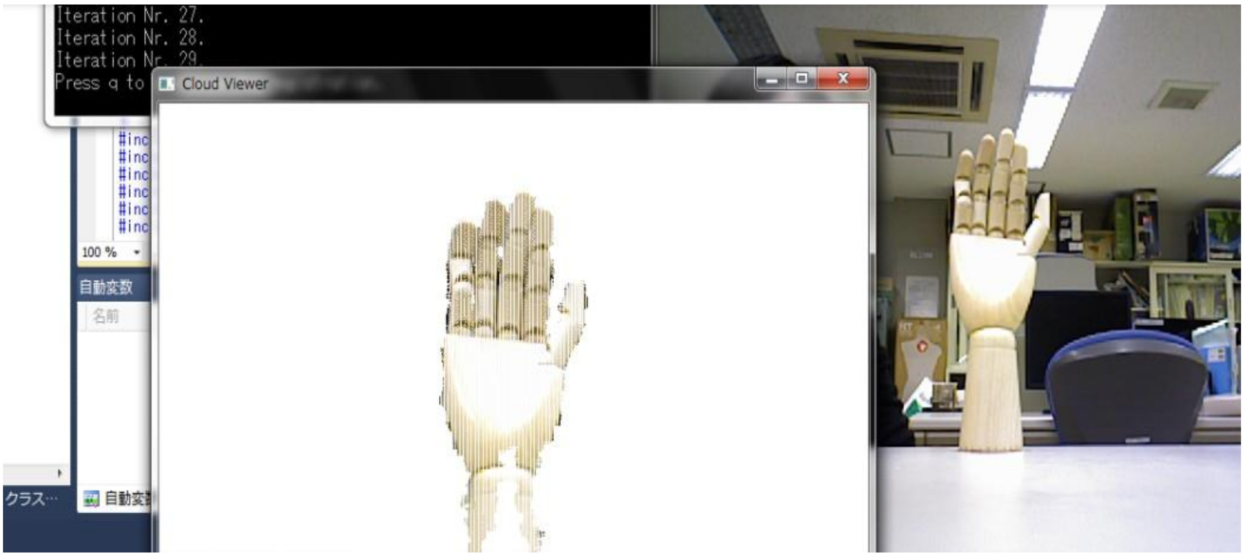

Figure.6 Recognition system of finger gesture

5. Discussions

We used 3D camera to utilize in evaluating the sign language recognition system. 3D sensor is useful for developing sign-language recognition system because extracting a hand shape was very easy. Until now, the recognition rate of the finger alphabets was about 40%, and low recognition rate was the biggest problem. Through the method of using two Kinects, we found that we can get more datasets of hand than use one, from the section 3, it is fully proved that the data of superposition can construct hand gesture more consummately. The most important part is the step of 3D processing using PCL, because although the Kinect can just take the part which we need, it still left so many noises, outliers and so on. However PCL have a strong power to deal with the trouble, it is introduced through section 4.

6. Conclusions

In conclusion, we proposed to develop a recognition system of sign language with a 3D sensor. To date, we had done the important first step, generating 3D data from the sensor with the least noise. Until now, although Kinect had been used to research about the part which are related to sign language [6], this kind of method has not been reported. The previous research still takes us very important messages and it is a very important part for us to research the recognition system of sign language.

About the future research, the most main work is evaluating the system of sign language. Although the previous job had been finished as it was expected, it is not sure whether the method and condition match the JSL system until we get the data of rate of recognition. Evaluation the datasets which had not been processed and which had been can prove the significance of the proposed approach, so we will use some ways to evaluate the rate of the recognition. Through the result of experiment we forecast that there will be some problems appeared, so the second mission is to solve and discuss about the miss, and perfect the JSL recognition system in further study.

7. References and Citations

[1]*Kinect Official Web Site* (2012) Available at <http://www.xbox.com/kinect/>.[Accessed 21 July 2012]

[2] *Kinect* (2011). Available at <http://en.wikipedia.org/w/index> [Accessed 21 June 2011]

[3] Lu Xia, Chia-Chih Chen and J. K. (2011) *Human Detection Using Depth Information by Kinect* [Online PDF].Available at <http://www.nattee.net/sites/default/files/Human%20Detection%20Using%20Depth%20 Information%20by%20Kinect.pdf > [Accessed 20-25 June 2011]

[4] Iason, Oikonomidis and Nikolaos, Kyriazis and Antonis A. Argyros. (2011)*Efficient Model-based 3D tracking of Hand Articulations using Kinect* [Online PDF]. Available at<http://www.ics.forth.gr/_ Publications /2011_09_bmvc_kinect_hand_tracking.pdf> [Accessed 29 August - 2 September 2011]

[5] K. Khoshelham. (2011) *Accuracy Anal Ysis of Kinect Depth Data* [Online PDF]. Available at< http://www. int-arch-photogramm-remote-sens-spatial-inf-sci.net/XXXVIII-5-W12/133/2011/isprsarchives-XXXVIII-5-W12-133-2011.pdf > [Accessed 29-31 August 2011]

[6] Matthew Tang. (2011) *Recognizing Hand Gestures with Microsoft's Kinect* [Online PDF].Available at < http://www.stanford.edu/class/ee368/Project_11/Reports/Tang_Hand_Gesture_Recognition.pdf > [Accessed 29-31 August 2011]

[7] Christian Vogler, Siome Goldenstein(2007)*Facial movement analysis in ASL*[Online PDF]. Available at< http://link.springer.com/article/10.1007%2Fs10209-007-0096-6?LI=true#page-1 >[Accessed 29-31 May 2012]